First passage time of N excluded-volume particles on a line

Igor M. Sokolov*
Institut für Physik, Humboldt Universität zu Berlin, Newtonstraße 15, 12489 Berlin, Germany

Ralf Metzler†
NORDITA, Nordic Institute for Theoretical Physics, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark

Kiran Pant and Mark C. Williams‡
Department of Physics, Northeastern University, 111 Dana Research Center, Boston, Massachusetts 02115, USA
(Received 12 June 2005; published 17 October 2005)

Motivated by recent single-molecule studies of proteins sliding on a DNA molecule, we explore the targeting dynamics of N particles ("proteins") sliding diffusively along a line ("DNA") in search of their target site (specific target sequence). At lower particle densities, one observes an expected reduction of the mean first passage time proportional to N^{-2}, with corrections at higher concentrations. We explicitly take adsorption and desorption effects, to and from the DNA, into account. For this general case, we also consider finite-size effects when the continuum approximation based on the number density of particles breaks down. Moreover, we address the first-passage-time problem of a tagged particle diffusing among other particles.

DOI: 10.1103/PhysRevE.72.041102 PACS number(s): 05.40.−a, 87.15.Vv, 02.50.−r

I. INTRODUCTION

DNA-binding proteins can either be bound specifically—i.e., such that the structure of the bound proteins exactly matches the entire DNA sequence it covers, involving Gibbs free energies of some 10 kcal/mol and above—or they can be bound nonspecifically with lower Gibbs free energies. Nonspecific binding occurs when the bound protein matches only part of the covered DNA sequence. A recent study showed that the repressor protein in λ-infected E. coli bacteria is bound nonspecifically with a Gibbs free energy of some 4 kcal/mol, causing under typical conditions nearly 90% of the repressor proteins to be bound nonspecifically [1]. In such a weak binding state, the protein can slide along the DNA, performing a one-dimensional (1D) diffusion process.

One of the primary tasks of DNA-binding proteins is the regulation of gene expression—i.e., to determine whether (or not) a certain gene on the genome is going to be transcribed by RNA polymerase. Having such processes in mind, we refer to these binding proteins as transcription factors (TF's) in what follows. The typical target search time of such a TF has received renewed attention [2–5] after the detailed investigations by Berg, Winter, and von Hippel [6]. One-dimensional sliding motion of DNA-binding proteins along the DNA molecule is an important ingredient in addition to three-dimensional volume diffusion in the efficient specific target search that is observed in experiments [6–8]. There exist, however, situations when the complete target search process of DNA-binding proteins occurs while being nonspecifically attached to the DNA molecule—i.e., without detaching from the DNA before hitting the target. This could be recently proved for bacteriophage T4 single-stranded DNA binding protein gp32 [9,10].

Gene regulation is a highly relevant example of a first passage time process, which can, in addition, be probed experimentally on the single-molecule level. While one usually considers the first passage of a single random walker or an ensemble of phantom random walkers, the sliding proteins on the DNA are clearly mutually excluding. To understand their target search quantitatively, one needs a theoretical model for the first passage of nonphantom particles. Surprisingly, there have been studied only a few cases of diffusion processes of mutually excluding particles—for instance, the diffusivity of particles on a line [11]. It should be noted that while some of the results below are known *per se* for the case of one-particle diffusion or for phantom particles [12,13], in the present case they are based on a mapping of the case of impenetrable particles, a problem that, to our knowledge, has not been studied so far. We also note that the problem pursued here is therefore also of a more generic interest, pertaining to the modeling of charge carrier motion in effectively one-dimensional geometries (nanowires, etc.) or traffic flow, among others.

In what follows, we establish a theory for the first passage dynamics of mutually excluding particles along a line ("DNA"). We explicitly take adsorption of particles to and desorption from the DNA into account, mimicking possible volume excursions of the proteins. Apart from the dilute case, we also address the dense case and the possibility of having more than one species of particles. Our analytical findings are corroborated by simulations.

II. SCALING APPROACH

In the simplest case when N identical, mutually excluding particles of size λ diffuse along a line of length L, we can

*Electronic address: igor.sokolov@physik.hu-berlin.de
†Electronic address: metz@nordita.dk
‡Electronic address: mark@neu.edu

obtain insight into the associated first passage process from scaling arguments. To be precise, the first passage is considered for a target placed at the origin ($x=0$) for particles that are initially randomly distributed along the line L. We first address the dilute case when the length $N\lambda$ occupied by the sliding particles can be neglected in comparison to the length L. Finite-size effects are regarded at the end of this section.

On average, it takes a random walker the time $T \simeq L^2/D_{1D}$ to cover the distance L by unbiased diffusion. Here, D_{1D} is the diffusion coefficient for 1D motion on a line and the symbol $\simeq$ indicates that we neglect constant prefactors. If there are N identical particles placed randomly over the line L, they are separated by an average length L/N; i.e., each of them has a free diffusion length L/N. For the first of these particle to hit the target site, this requires a characteristic target search time

$$T_{\mathrm{dil}}(N) \simeq \frac{L^2}{D_{1D}N^2} = \frac{1}{D_{1D}n_0^2}, \tag{1}$$

with $n_0=N/L$ being the line density of particles with dimension 1/length. The index is meant to distinguish the dilute result from the result (3) when finite-size effects come into play.

We performed a simulation of particles on a line during which each particle attempts a jump to its left or right nearest-neighbor lattice point per unit time. In case the corresponding site is occupied, the step is forbidden and the particle remains at its original site. The associated (dimensionless) diffusion coefficient of a *single* particle per unit length square and unit time is $D_{1D}=1/2$. Figure 1 shows the results for the mean target search time $T(n_0)$ of this simulation in dependence of the density n_0 of particles. We find nice agreement with the expected inverse-square dependence of $T(n_0)$ on the density n_0. The line through the data points corresponds to the analytical result from Eq. (1) with a prefactor given by Eq. (17) without adjustable parameters. The results demonstrate that the theoretical approximation leading to the $1/N^2$ behavior remains reasonable even at rather high concentrations, at which the interparticle distance becomes of the order of the step lengths. In the next section, we derive the $1/N^2$ scaling analytically in a continuum approximation.

Experimentally—for instance, in *in vivo* studies of proteins binding to a DNA molecule—the diluteness condition is perfectly adequate; compare, for instance, Ref. [14]. By increasing the protein concentration or their binding strength through different ambient salt conditions, the concentration of bound proteins can be increased such that finite-size effects indeed come into play. Similarly, the presence of many different species of proteins leads to a rather crowded DNA molecule. Similar considerations apply, of course, to other systems. Defining the occupation ratio

$$f = \frac{N\lambda}{L}, \tag{2}$$

we can express the diluteness condition through $f \ll 1$. To include finite-size effects when this limit is not fulfilled

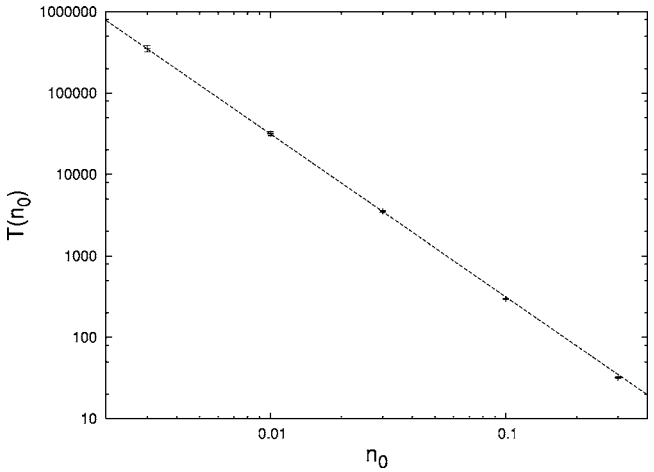

FIG. 1. Mean first passage time $T(n_0)$ in a one-sided system, as a function of the density n_0 of excluding walkers that cannot occupy the same lattice site. The target is placed at $x=0$. The maximum density is $n_0=30\%$, whereby each particle occupies one lattice site. The dashed line corresponds to the exact result $T(n_0)=\pi/n_0^2$ from Eq. (17), with the dimensionless diffusion coefficient $D_{1D}=1/2$. The simulation data agree nicely with the dilute limit, with only a slight deviation for larger densities. Each data point corresponds to 10^5 runs, except for 10^3 realizations for the lowest density. Note the comparatively small error bars.

in our scaling approach, we only need to consider the reduced length of the line available to the random walking particles. This reduced length is $L_{\mathrm{red}}=L-N\lambda$, so that we obtain

$$T(N) \simeq \frac{(L-N\lambda)^2}{D_{1D}N^2} = T_{\mathrm{dil}}(N)(1-f)^2, \tag{3}$$

for the scaling of the mean target search time with the number N of particles. Figure 2 compares the dilute $1/N^2$ scaling

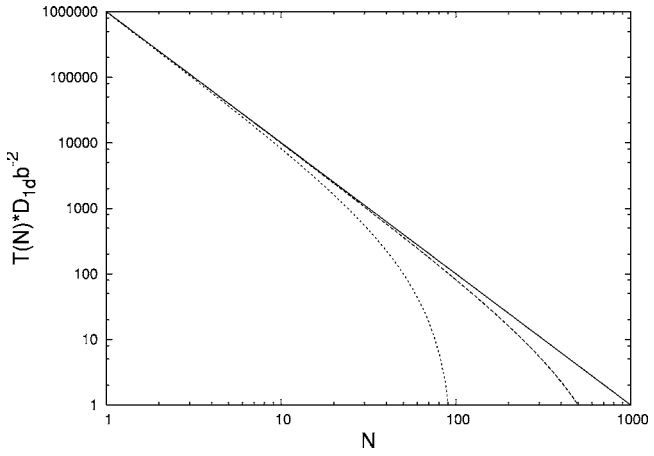

FIG. 2. Behavior of the mean first passage time $T(N)$ as a function of the number N of TF's attached to a DNA of length 1000 according to Eq. (3), for the dilute case (solid line), TF size $\lambda=1$ (long-dashed line), and $\lambda=10$ (short-dashed line). Excluded-volume effects reduce the target search time $T(N)$.

with the finite-size effects predicted by the excluded-volume expression (3).

III. CONTINUUM APPROXIMATION

In this section, we verify the above scaling result for the dilute case $T(N) \simeq L^2/(D_{1D}N^2)$ through an analytic treatment in the continuum approximation, replacing the individual TF's through the particle density $n(x,t)$. In addition, we include explicitly adsorption and desorption effects with rates k_0 and k_1.

To be able to take the continuum limit, we consider large systems (long DNA) with many ($N \gg 1$) searching TF's, such that the concentration of TF's on the DNA is much smaller than unity; that is, $f \ll 1$. In other words, the diffusion time through the whole system, $T_1 \simeq L^2/D_{1D}$, is much larger than the typical first passage time corresponding to the characteristic target search time, being of the order of $T \simeq 1/(f^2 D_{1D})$. We note that for $f \ll 1$, the fraction f depends linearly on the volume concentration C of TF's, according to the McGhee–von Hippel isotherm [15].

We start by considering a one-sided problem (one target site at $x=0$ of a semi-infinite DNA). The time evolution of the number concentration $n(x,t)$ at position x at time t on the semi-infinite interval is then given by the diffusion-reaction equation

$$\frac{\partial n}{\partial t} = D_{1D}\frac{\partial^2}{\partial x^2}n - k_1 n + k_0. \tag{4}$$

Apart from diffusion, in this equation we take into account adsorption (with rate k_0) and desorption (with rate k_1) of the TF's, where the desorption is proportional to the number concentration of TF's on the DNA. Apart from real physical absorption-desorption processes, this approach might mimic other nonlocal processes such as macrohops (3D volume sojourns) and intersegmental transfer (hopping from one segment of the DNA to another, chemically remote segment—that is, close by in geometric space due to looping of the DNA) in a mean-field sense. Following Smoluchowski's approach to diffusion-controlled reactions, we represent the target site by an absorbing boundary condition at $x=0$; i.e., when a diffusing particle hits this site, it will be removed. The possibility of double occupation of sites is disregarded, as it represents a higher-order effect proportional to f^2. Moreover, the fact that particles are impenetrable to each other does not change the behavior at low concentrations, since, neglecting the excluded volume, on encounter of two particles it does not matter whether the right particle always stays to the right of the other particle (impenetrable particles) or whether they change roles and the right particle becomes the left one (phantom particles), as long as the particles are indistinguishable, in contrast to the case of distinguishable particles addressed below. Finite-size effects due to high occupation, violating the diluteness condition $f \ll 1$, will also be addressed below.

Finding the target corresponds to the event when the first particle hits the target site. Mathematically, this is equivalent to the first passage time of a particle from a site $x>0$ to $x=0$, given by the particle flux into the reaction center, $j(t) = D_{1D}\partial n/\partial x|_{x=0}$. The survival probability $\mathcal{S}(t)$ of the target site (i.e., the probability of not yet having been hit by a TF, not to be confused with the survival of the particles along the DNA) is consequently given by the first-order kinetic equation

$$\frac{d}{dt}\mathcal{S}(t) = -j(t)\mathcal{S}(t). \tag{5}$$

The change of the survival probability, of not having been hit, of the target site is thus the product of the probability of not having been hit previously times the magnitude of the influx of particles. The formal solution of Eq. (5) reads

$$\mathcal{S}(t) = \exp\left(-\int_0^t j(t')dt'\right). \tag{6}$$

In what follows we use the notation $J(t) = \int_0^t j(t')dt'$. The first passage time *density* is then given by

$$\psi(t) = -\frac{d}{dt}\mathcal{S}(t) = j(t)\exp[-J(t)]. \tag{7}$$

In our one-sided problem, the mean first passage time becomes $T = \int_0^\infty t\psi(t)dt = -\int_0^\infty t[d\mathcal{S}(t)/dt]dt$—i.e.,

$$T = \int_0^\infty \mathcal{S}(t')dt'. \tag{8}$$

To obtain an explicit expression for $\mathcal{S}(t)$, we solve the reaction-diffusion equation (4) by Laplace transformation techniques. With the initial condition $n(x,0) = n_0\Theta(x)$, where $\Theta(x)$ is the Heaviside jump function, we obtain for all $x>0$ for the Laplace transform $\tilde{n}(x,u)$:

$$u\tilde{n} - n_0 = D_{1D}\frac{\partial^2}{\partial x^2}\tilde{n} + \frac{k_0}{u} - k_1\tilde{n}, \tag{9}$$

i.e., a linear inhomogeneous differential equation of the form

$$\tilde{n}'' - \Lambda\tilde{n} + B = 0, \tag{10}$$

with $\Lambda = (k_1+u)/D > 0$ and $B = (k_0/u + n_0)/D > 0$. The boundary conditions we impose are of the absorbing Dirichlet type $n(0,u)=0$ at the target site placed at the origin and the natural boundary condition $n(x,u) < \infty$ for $x \to \infty$. The corresponding solution reads

$$\tilde{n}(x,u) = \frac{k_0 + un_0}{u(k_1+u)}\left(1 - e^{-x\sqrt{(k_1+u)/D_{1D}}}\right). \tag{11}$$

From this expression, we find, for the flux $j(t)$ in Laplace space,

$$\tilde{j}(u) = D_{1D}\left.\frac{\partial\tilde{n}(x,u)}{\partial x}\right|_{x=0} = \sqrt{D_{1D}}\frac{k_0 + un_0}{u\sqrt{k_1+u}}, \tag{12}$$

an expression whose inverse Laplace transform can be calculated explicitly, yielding

$$j(t) = \sqrt{D_{1D}}\left[\frac{k_0}{\sqrt{k_1}}\,\mathrm{erf}\,\sqrt{k_1 t} + n_0 \frac{e^{-k_1 t}}{\sqrt{\pi t}}\right]. \tag{13}$$

The survival probability of the target site then is given by $\mathcal{S}(t)=\exp[-J(t)]$ with

$$J(t) = \sqrt{D_{1D}}\left[\frac{k_0}{k_1}\left(t\sqrt{k_1}\,\mathrm{erf}\,\sqrt{k_1 t} - \frac{\mathrm{erf}\,\sqrt{k_1 t}}{2\sqrt{k_1}} + \frac{\sqrt{t}}{\sqrt{\pi}}e^{-k_1 t}\right) + n_0 \frac{\mathrm{erf}\,\sqrt{k_1 t}}{\sqrt{k_1}}\right]. \tag{14}$$

Without adsorption and desorption (i.e., $k_0=k_1=0$), we obtain the survival provability

$$\mathcal{S}(t) = \exp\left(-2n_0\sqrt{\frac{D_{1D}t}{\pi}}\right) \tag{15}$$

and first-passage-time density

$$\psi(t) = \frac{n_0\sqrt{D_{1D}}}{\sqrt{\pi t}}\exp\left(-2n_0\sqrt{\frac{D_{1D}t}{\pi}}\right). \tag{16}$$

We thus find for the mean first passage time $T=\int_0^\infty \mathcal{S}(t)dt$ the simple form

$$T_{\mathrm{line}} = \frac{\pi}{2}\frac{1}{n_0^2 D_{1D}}, \tag{17}$$

showing the typical n_0^{-2} dependence on the initial concentration.

The survival provability for the general case with nonvanishing rates k_0 and k_1 becomes

$$\mathcal{S}(t) = \exp\left[-\sqrt{D_{1D}}(k_0 k_1 t - k_0/2 + n_0 k_1)\frac{\mathrm{erf}\,\sqrt{k_1 t}}{k_1^{3/2}} - \frac{k_0}{k_1}\frac{\sqrt{D_{1D}t}}{\sqrt{\pi}}\exp(-k_1 t)\right]. \tag{18}$$

Eventually (for $t \gg k_1$), an exponential decay $\sim\exp(-\sqrt{D_{1D}}k_0 k_1 t)$ is reached. From this asymptotic behavior, we can deduce the approximate dependence $T\approx(\sqrt{D_{1D}}k_0 k_1)^{-1}$. As the adsorption rate k_0 is proportional to the concentration C of TF's in volume, we obtain the typical $T\sim C^{-1}$ dependence of the mean target search time under volume exchange conditions. This contrasts the $T\sim n_0^{-2}$ behavior for 1D sliding exchange found in Eq. (17). Given that $n_0 \simeq C$ for $n_0 \ll 1$, the latter corresponds to the $T\simeq C^{-2}$ scaling demonstrated in Fig. 5 below. In general, there will be a combination of both behaviors, depending on the values of the various system parameters.

In the case of no adsorption $k_0=0$ but nonvanishing desorption $k_1\neq 0$, which corresponds to a situation with vanishing concentration of TF's in the free volume, the function

$$J(t) = \sqrt{D_{1D}}n_0\frac{\mathrm{erf}\,\sqrt{k_1 t}}{k_1^{1/2}} \tag{19}$$

is bounded from above, by $n_0\sqrt{D_{1D}/k_1}$ and the survival probability $\mathcal{S}(t)$ never reaches zero (all particles desorb with a nonzero probability without ever reaching the target site $x=0$), and the probability density $\psi(t)$ is a nonproper one, corresponding to a diverging mean first passage time. In all other cases $\psi(t)$ is a proper probability density, and the mean target search time T is finite.

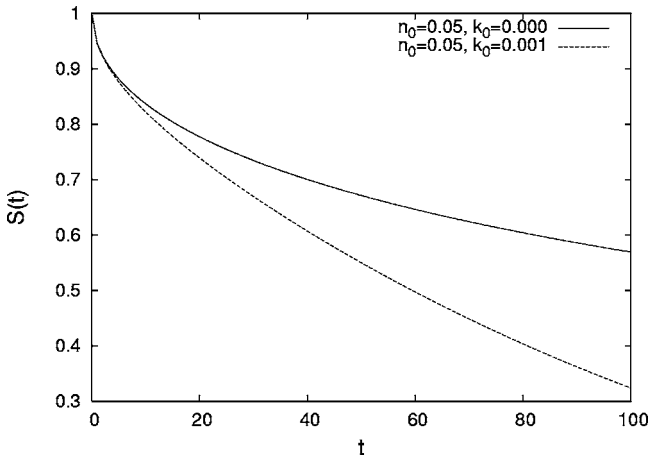

FIG. 3. Survival probability of the target site—i.e., the probability that no TF has reached the specific binding site—according to Eqs. (15) and (21). This case corresponds to vanishing desorption, $k_1=0$. The plot parameters are indicated in the figure.

Performing an expansion in powers of t (the corresponding series contains only the half-integer powers), we find for the function $J(t)$ in the general case with finite k_0, k_1:

$$J(t) = \sqrt{\frac{D_{1D}}{\pi}}\left[2n_0 t^{1/2} + \frac{2}{3}k_1\left(2\frac{k_0}{k_1} - n_0\right)t^{3/2} + \frac{1}{15}k_1^2\left(-4\frac{k_0}{k_1} + 3n_0\right)t^{5/2} + \cdots\right], \tag{20}$$

so that the nth term of the expansion has a structure $k_1^{n-1}(a_n k_0/k_1 + b_n n_0)t^{(2n-1)/2}$. Thus, in essence, this expansion corresponds to an expansion in powers of k_1. Note that $k_0/k_1=n_s$ is a steady-state concentration of proteins in the absence of the absorbing target site. As long as both k_0 and k_1 are small, the overall behavior given by Eq. (17) is preserved, provided the initial concentration n_0 is not too small. In the case without desorption ($k_1\to 0$) we get

$$\mathcal{S}(t) = \exp\left(-2n_0\sqrt{\frac{D_{1D}t}{\pi}} - \frac{4}{3}\sqrt{\frac{D_{1D}}{\pi}}k_0 t^{3/2}\right). \tag{21}$$

This equation is important in what follows, when finite-size effects are considered.

In Fig. 3, we plot the survival probabilities from Eqs. (15) and (21) for an initial line density of TF's of $n_0=0.05$. Both cases correspond to vanishing desorption rate, $k_1=0$, and therefore $\mathcal{S}(t)$ decays completely for large times. This decay of the survival probability of the target site in both cases follows the same behavior for short times, until the adsorption according to Eq. (21) leads to a faster target search and therefore to a quicker decay of $\mathcal{S}(t)$. Similarly, Fig. 4 shows the survival probabilities in the general case corresponding to Eq. (18); note the logarithmic ordinate. For vanishing adsorption but finite desorption, the expected incomplete decay

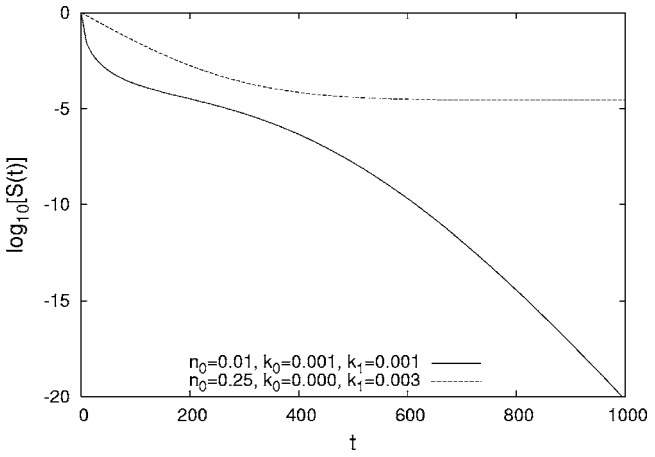

FIG. 4. Survival probability $\mathcal{S}(t)$ from Eq. (18) for finite desorption rate k_1. Note the logarithmic ordinate; the plot parameters are indicated in the figure. The incomplete decay in the case of vanishing adsorption, $k_0=0$, is distinct.

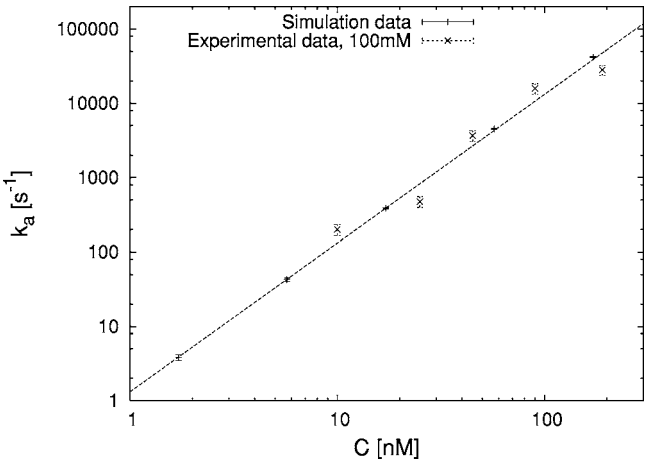

FIG. 5. Dimensional binding rate k_a in 1/s as a function of protein concentration C in nM, converted from Fig. 1 for parameters corresponding to 100 mM salt. The fitted 1D diffusion constant for sliding along the dsDNA is $D_{1D}=3.3\times10^{-9}$ cm^2/sec, located nicely within the experimental value $10^{-8}-10^{-9}$ cm^2/sec [9].

of $\mathcal{S}(t)$ is observed, whereas for finite adsorption and desorption the transition between the different contributions in expression (18) is visible, eventually approaching the simple exponential pattern, which corresponds to a straight line in this plot.

The two-sided problem (a ring geometry with a perimeter that is much larger than the typical interparticle distance) corresponds to the situation where two competing processes occur; i.e., the survival probability of having an empty target site changes in time through the influx of TF's from both sides. This practically corresponds to using twice the probability current j in Eq. (5) due to symmetry and therefore to

$$\mathcal{S}(t) = \exp[-2J(t)], \tag{22}$$

with $J(t)$ given by Eq. (14). The corresponding mean first passage time for the case $k_0=k_1=0$ is then given by

$$T_{\text{ring}} = \frac{\pi}{8}\frac{1}{n_0^2 D_{1D}}, \tag{23}$$

which is by a factor of 4 smaller than in the one-sided case. The result (23) is also confirmed by numerical simulations. We note that the reduction by a factor of 4 can be easily understood by mapping the circle with one absorbing site onto a line whose both ends are absorbing boundaries. It then corresponds to two one-sided geometries as considered above, but with an effective length of $L/2$. With the definition of the initial number concentration n_0, this reproduces the factor of 4.

For direct comparison with the experimental data, Fig. 5 shows an alternative way to present the numerical data from Fig. 1 in dimensional form of the rate k_a in units of 1/s versus the volume protein concentration C in units of M. For the conversion, we use the relation $n_0=K_{\text{ns}}\mu C$, with the nonspecific binding constant $K_{\text{ns}}=2.5\times10^5$ M^{-1} and the SSB binding size $\mu=7$ in units of nucleotides [9,16]. By a logarithmic least-squares fit to the shown data measured at 100 mM salt, we obtain for the dimensional diffusion constant D_{1D} of 1D sliding along the dsDNA the value $D_{1D}=3.3\times10^{-9}$ cm^2/sec, which is nicely within the reported range $10^{-8}-10^{-9}$ cm^2/sec for this salt concentration [9,16]. This corroborates the validity of our rather simple analytical model for the target search of a truncate of the gp32 protein. Note that the experimental situation with two target sites at either end of the DNA molecule corresponds to the result (23).

IV. FINITE-SIZE EFFECTS

In the previous section we discussed the case of a semi-infinite DNA and argued that the case when the target site is situated somewhere in the middle of the molecule can be inferred from that result.

Now we consider the finite-size situation (again one sided), with a target site situated on one side of a chain and with another side closed by a "stopper"—for instance, a polystyrene bead in an optical tweezers setup—such that the sliding proteins observe a reflecting boundary condition. This consideration is necessary to discuss finite-size effects and also to derive results explicitly used in the next section. The situation where there are two portions of the chain to the left and to the right from the target site corresponds to two independent reaction channels, so that the mean reaction time follows from those in the left and in the right intervals: $1/T=1/T_L+1/T_R$.

To consider this situation on an interval of length L with exactly N TF's, we have to solve our equation (4) with the boundary conditions $n(0,t)=0$ (reacting center) and $n'(L,t)=0$ (blocked end) and with initial condition $n(x,0)=n_0=N/L$. Under Laplace transformation this leads us to Eq. (10), now with corresponding boundary conditions. The solution then becomes

$$\tilde{n}(x,u) = \frac{b}{\lambda}(1-\cosh x\sqrt{\lambda}+\tanh L\sqrt{\lambda}\,\sinh x\sqrt{\lambda}), \tag{24}$$

so that the Laplace transform of the probability current reads

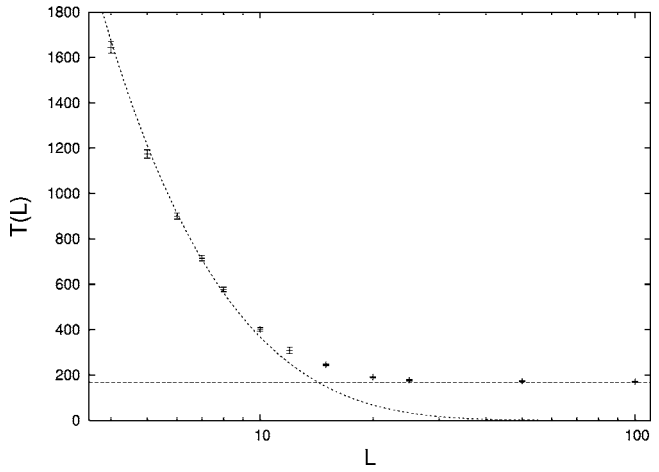

FIG. 6. Mean first passage time $T(L)$ of target search in a pronouncedly finite system of length L, corresponding to the one-sided system with target site at $x=0$ and reflecting boundary condition at $x=L$. We chose the parameters $D_{1D}=1/2$ for the dimensionless diffusion constant, the initial protein density $n_0=10\%$, the adsorption rate $k_0=10^{-4}$, and vanishing desorption rate $k_1=0$. The curved dashed line corresponds to Eq. (30), which approximates small systems, while the horizontal dashed line at $T=168$ is determined by numerical integration of $\mathcal{S}(t)$, Eq. (21). Each data point represents 10^5 runs. Again, note the rather small error bars.

$$\tilde{j}(u) = \sqrt{D_{1D}} \frac{k_0 + u n_0}{u\sqrt{k_1 + u}} \tanh \frac{L\sqrt{k_1 + u}}{\sqrt{D_{1D}}}. \tag{25}$$

This expression tends to our equation (12) in the limiting case $L \to \infty$. In the general case there does not exist a closed expression for $j(t)$ and thus for T. However, for small enough $L\,(L\sqrt{k_1})/\sqrt{D_{1D}} \ll 1$—i.e., in the case when the diffusion time along the interval L is so small that practically no desorption takes place) we can approximate $\tanh x$ by the value of its argument and obtain

$$\tilde{j}(u) = \frac{L}{u}(k_0 + n_0 u). \tag{26}$$

This result implies

$$j(t) \simeq [k_0 + n_0 u]L \tag{27}$$

and

$$J(t) \simeq [k_0 + n_0 \delta(t)]L, \tag{28}$$

so that we find the survival probability

$$\mathcal{S}(t) \simeq e^{-n_0 L} e^{-k_0 L t}. \tag{29}$$

The latter result leads to the approximate form

$$T \simeq \frac{e^{-n_0 L}}{k_0 L}. \tag{30}$$

This behavior will be of importance in what follows. In Fig. 6, we show results from simulations on a finite system, demonstrating that the predicted asymptotic behavior, Eq. (30), in fact describes the behavior of the system quite accurately for smaller L and eventually reaches a constant value for larger system sizes (note that the density n_0 of proteins is kept constant).

The enumerator in Eq. (30) is the probability that no TF's are initially present in the interval. If there were any, the target finding could typically occur within the time interval $\tau \simeq L^2/(N^2 D_{1D})$. However, there is a nonzero probability (equal to $e^{-n_0 L}$) that no TF's are initially found in the interval. In this case one has to wait on the average $1/k_0 L$ until a TF is adsorbed, a slow process which governs the overall expression kinetics. This is the true asymptotics of the waiting time in the case when k_0 is very small, so that the absorption time $1/k_0 L$ is much larger than the typical diffusion time over the interval L, being of the order of $T_D \simeq L^2/D_{1D}$.

We note that the situation considered here is pertinent to the grand canonical ensemble (N fluctuates around the mean value $N=n_0 L$); the canonical situation (N fixed) is discussed in the Appendix.

V. DIFFERENT SPECIES OF TRANSCRIPTION FACTORS

The picture changes if we regard TF's of different species. If the relative concentration of "relevant" TF's is high enough, the situation stays practically the same as before, since the "dummy" proteins simply act as the effective "boundaries" reducing the length of the search region to $\tilde{L}=L/N_{\text{dummy}}$ around the target site. This simple assumption is realistic since specifically bound TF's would, for most practical purposes, represent immobile barriers (the Gibbs free energy for specific binding is larger than for nonspecific binding). Since, however, the effective search time depends only on the overall concentration of relevant proteins, the typical search time will not change considerably, unless the situation occurs that no relevant proteins are encountered within the search region with appreciable probability. This situation takes place if the concentration of dummy TF's becomes of the order of or larger than the concentration of relevant TF's. The reaction can take place only if a relevant particle is situated in the same interval between the barriers as the target site is. The mean waiting time in this case can be obtained from the result of the previous section. Let us consider the interval to one side of the target site. The length of this interval be L. Assuming independent positions of all TF's, we can obtain the joint probability distribution of the length of the interval between the reaction center and the next boundary protein and of the mean initial concentration of relevant TF's inside, $p(n_0, L)$ (note here that the variable N is discrete, while L is continuous). Noting that the actual initial concentration n_0 in each realization is $n_0=N/L$ and that the density of the waiting time distribution for a given (nonfluctuating) n_0 and L is given by a function $\psi(t; L, n_0, k_0, k_1)$, the overall waiting time distribution yields as a mixture—i.e., by simple averaging,

$$\Psi(t) = \int_0^\infty dL p(L) \psi(t; L, n_0, k_0, k_1), \tag{31}$$

where $p(L)$ is the probability density to find a specifically bound TF at the distance L from the target site. The corresponding mean waiting time in the one-sided problem follows then as

$$\bar{T} = \int_0^\infty dL p(L) T(L, N/L, k_0, k_1), \tag{32}$$

with the weight $T(L,n_0,k_0,k_1)=\int_0^\infty t\psi(t;L,n_0,k_0,k_1)dt = \int_0^\infty \mathcal{S}(t;L,n_0,k_0,k_1)dt$, where ψ represents the waiting time probability density function and $\mathcal{S}(t)=\int_t^\infty \psi(t')dt'$ being the survival probability. Assuming Poissonian statistics of the distribution of TF's we find that the distribution of L is exponential, $p(L)=c_s e^{-c_s L}$, with c_s being the concentration of specifically bound TF's. Here a clear difference between the one-sided and two-sided problems emerges.

We give here the explicit results only for the case when $k_0 < c_s^3 D$. The mean waiting time in a one-sided problem is then given by Eq. (30). Averaging this expression over the distribution of the lengths of the intervals between the target site and the blocking by specifically bound TF we see that the corresponding expression

$$\bar{T}_1 = \frac{c_s}{k_0}\int_0^\infty \frac{1}{L} e^{-(n_0+c_s)L} dL \tag{33}$$

diverges. This divergence has to do with the possibility of immediate blocking, which becomes evident when we return to the initial, discrete situation: since it is possible that a specifically bound TF is an immediate neighbor of the target site, the reaction is simply impossible. Of course, one can overcome this difficulty by assuming that there exists a minimal size of such an interval $L_{\min}$ (or that the immediate absorption of a relevant TF on the center is possible, which, from the mathematical point of view, is equivalent to putting this minimal length equal to a size of the target site). Assuming this $L_{\min}$ to be small compared to all other spatial scales of the problem so that $p(L) \simeq c_s e^{-c_s L}/e^{-c_s L_{\min}}$, we obtain, asymptotically,

$$\begin{aligned}\bar{T} &= \frac{c_s}{k_0 e^{-c_s L_{\min}}}\int_{L_{\min}}^\infty \frac{1}{L} e^{-(n_0+c_s)L} dL \\ &= \frac{c_s}{k_0 e^{-c_s L_{\min}}}\Gamma[0,(n_0+c_s)L_{\min}]\end{aligned} \tag{34}$$

[$\Gamma(x,y)$ being the incomplete Γ function], which grows very slowly (logarithmically) for $L_{\min}\to 0$.

Let us now turn to the two-sided situation. In this case the survival probability of the target site $\Psi(t)=\mathcal{S}(t;L_1,n_0,k_0,k_1)\mathcal{S}(t;L_2,n_0,k_0,k_1)$ where L_1 and L_2 are the lengths of free intervals to the left and to the right from the target site: it survives up to time t if the TF comes to it neither from the right nor from the left. Using Eq. (29) valid for k_1 small we get

$$\begin{aligned}\bar{T}_2 &= \int_0^\infty dt \int_0^\infty \int_0^\infty dL_1 dL_2 c_s^2 e^{-(c_s+n_0+k_0 t)(L_1+L_2)} \\ &= c_s^2 \int_0^\infty \frac{dt}{(c_s+n_0+k_0 t)^2} = \frac{c_s^2}{(c_s+n_0)k_0}.\end{aligned} \tag{35}$$

In this case the mean waiting time (still fully defined by the absorption) is finite, since the probability that the target site is blocked from both sides is negligibly small.

VI. CONCLUSIONS

We derived analytically the first passage time behavior for a set of N mutually excluding particles on a line in the dilute limit. As predicted from scaling arguments, the corresponding mean first passage time decays inversely with the square of the number of particles. The analytical behavior was corroborated by simulations results, showing nice agreement without a free parameter. Comparison with experimental results from the one-dimensional target search of the bacteriophage T4 protein gp32 produces a very reasonable fitted one-dimensional diffusion constant of the sliding protein.

Having in mind the target search of transcription factors on a long DNA, during which one-dimensional sliding motion along the DNA is interrupted by three-dimensional volume excursions, we included desorption from and adsorption to the DNA. These affect the time evolution of the survival probability of the specific target sequence, which may be of importance to the design of related *in vitro* experiments. Moreover, the obtained description may be relevant to other (bio)chemical systems as well as nanosetups—for instance, the one-dimensional diffusion of particles in a nanochannel and their escape through a T junction.

Finally, we discussed effects due to the finite size of the DNA (line) along which the diffusion takes place. This may be of importance for certain *in vitro* experiments employing a rather short stretch of DNA. The predicted behavior was corroborated (without adjustable parameter) by simulations. Similar effects arise when the first passage of an individual tagged particle is considered.

We note that our derivations were based on normal Markovian diffusion dynamics. To generalize our results to situations footing on long-tailed waiting time distributions, which cause a subdiffusive behavior, the standard procedure can be used to map the Markovian to the subordinated subdiffusive process [17] and the associated dynamical equation contains a fractional time derivative [18,19]. Intersegmental jumps [6] at places where by DNA looping chemically distant segments of the DNA come in close contact in physical three-dimensional space [20] can even give rise to Lévy flights [21,22]. The latter situation requires special care when comparing the efficiency between sliding motion along the DNA and the Lévy flight mixing under varied salt conditions, as explored in [23]. Another remark concerning our modeling in terms of the diffusion-controlled Smoluchowski picture is in order. Namely, in transport-controlled reactive systems, which are not overdamped, at shorter times there is the need to include the transient ballistic regime in the reaction scheme, as discussed in Ref. [24] starting from the Klein-Kramers picture. However, in our problem, the diffusion process is highly overdamped [25] and the Smoluchowski approach is therefore appropriate.

ACKNOWLEDGMENTS

We would like to thank Ulrich Gerland and Oleg Krichevsky for helpful discussions. This work was supported by Grant Nos. NIH GM 072462 and NSF MCB-0238190 and the Research Corporation.

APPENDIX

Here, we want to elucidate the role of finite-size effects and to stress the difference between the grand-canonical and canonical situations (i.e., when the number N of particles in the interval of length L is variable or fixed or prescribed by Poisson statistics, respectively). All our considerations in the main text were pertinent to the last situation corresponding to the grand-canonical ensemble, which seems to be experimentally relevant. Here, for completeness we discuss the other case.

We concentrate on the situation without adsorption-desorption processes ($k_0=k_1=0$). For noninteracting particles, the probability density $p(x,t)$ to find a particle at site x is described by the same equation (4), however now with the initial condition $p(x,0)=1/L$ corresponding to the normalization of the probability density. The overall survival probability of a given particle in the interval is simply given by $\Psi(t)=\int_0^L p(x,t)dx$. Performing integration over x in Eq. (24), giving now (for $n_0=1/L$, $k_0=k_1=0$) the Laplace-transformed $\tilde{p}(x,u)$:

$$\tilde{\Psi}(u)=\frac{1}{u}-\frac{\sqrt{D_{1D}}}{Lu^{3/2}}\tanh\left(\frac{L\sqrt{u}}{\sqrt{D_{1D}}}\right). \tag{A1}$$

For exactly N particles, the survival probability of the reaction center is $\mathcal{S}(t)=\Psi^N(t)$: it only survives if none of the particles arrived at it up to the time t and the mean survival time $T(N)=\int_0^\infty \mathcal{S}(t)dt=\int_0^\infty \Psi^N(t)dt$. For $N=0$ one has $\mathcal{S}(t)=1$, so that the mean waiting time diverges. For whatever finite N the mean waiting time is finite. It follows from the fact that the function $\Psi(t)$, which is non-negative and monotonously nongrowing, is integrable, and its time integral $T(1)=\lim_{u\to 0}\tilde{\Psi}(u)=L^2/3D_{1D}$. This means that for $t\to\infty$ this function decays faster than as t^{-1}, and thus its powers decay even faster and are integrable. For N small the value of $T(N)$ has to be calculated explicitly. For large N a simple asymptotic expression arises: In this case the mean waiting time is much smaller than L^2/D_{1D}, with small times corresponding to large $u\gg D_{1D}/L^2$. For such u one has $\tanh[\sqrt{D_{1D}}/(L\sqrt{u})]\to 1$ so that one can write

$$\tilde{\Psi}(u)\simeq =\frac{1}{u}-\frac{\sqrt{D_{1D}}}{Lu^{3/2}}. \tag{A2}$$

The inverse Laplace transform of this function gives us the small-t behavior of $\mathcal{S}(t)$—namely,

$$\Psi(t)\simeq 1-\frac{1}{L}\sqrt{\frac{4D_{1D}t}{\pi}}. \tag{A3}$$

For N large enough one then has

$$\begin{aligned}\mathcal{S}(t)=\Psi^N(t)&\simeq\left(1-\frac{1}{L}\sqrt{\frac{4D_{1D}t}{\pi}}\right)^N\\&\simeq\exp\left[N\ln\left(1-\frac{1}{L}\sqrt{\frac{4D_{1D}t}{\pi}}\right)\right]\\&\approx\exp\left(-\frac{N}{L}\sqrt{\frac{4D_{1D}t}{\pi}}\right),\end{aligned} \tag{A4}$$

which is exactly our equation (15) with $n_0=N/L$. The approximation is reasonably good starting from $N\sim 10$ particles. The canonical mean waiting time (i.e., the mean waiting time with *exactly* N particles in the interval) is then given by the same equation (17).

The grand-canonical result can be obtained from the canonical one by simply noting that the grand-canonical expression for $\mathcal{S}(t)$ corresponds to a weighted sum of the corresponding canonical waiting times:

$$\mathcal{S}(t)=\sum_N \Psi^N(t)p_N, \tag{A5}$$

where p_N is the probability to find exactly N particles within the interval. Taking this probabilities to follow a Poisson distribution, $p_N=[(n_0L)^N/N!]\exp(-n_0L)$, we get

$$\mathcal{S}(t)=\sum_{N=0}^{\infty}\frac{[n_0L\Psi(t)]^N}{N!}\exp(-n_0L)=\exp[n_0L(\Psi(t)-1)], \tag{A6}$$

which, for small t, again corresponds to our equation (15). Using an approximate short-time asymptotic expression for $\Psi(t)$, Eq. (A3), we again arrive at Eq. (A4). However, some care is required when interpreting this result.

Since $\Psi(t)$ is integrable and thus $\Psi(t)\to 0$ for $t\to\infty$, $\mathcal{S}(t)$ tends for $t\to\infty$ to a constant value, $\mathcal{S}(t)\to\exp(-n_0L)$, which is exactly the probability to have no TF's in the interval. In this case, of course, no reaction takes place at all. The mean waiting time, being the time integral of $\mathcal{S}(t)$, clearly, diverges. If we separate the first, constant, term in Eq. (A6), we get an expression for $\mathcal{J}(t)$ which (for $n_0L\gg 1$) is asymptotically the same as Eq. (A4), and calculating the mean waiting time, we again arrive at Eq. (17). This mean waiting time has, however, to be interpreted as the mean waiting time for the reaction, provided that it happens at all. The probability that it never happens is equal to $\exp(-n_0L)$ and is small but finite.

[1] A. Bakk and R. Metzler, FEBS Lett. **563**, 66 (2004); J. Theor. Biol. **231**, 525 (2004).
[2] M. Slutsky and L. A. Mirny, Biophys. J. **87**, 4021 (2004).
[3] M. Coppey, O. Benichou, R. Voituriez, and M. Moreau, Biophys. J. **87**, 1640 (2004).
[4] S. E. Halford and J. F. Marko, Nucleic Acids Res. **32**, 3040 (2003).
[5] U. Gerland, J. D. Moroz, and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **99**, 12015 (2002).
[6] O. G. Berg, R. B. Winter, and P. H. von Hippel, Biochemistry

20, 6929 (1981); R. B. Winter, O. G. Berg, and P. H. von Hippel, *ibid.* **20**, 6961 (1981).
[7] G. Adam and M. Delbrück, in *Structural Chemistry and Molecular Biology*, edited by A. Rich and N. Davidson (W. H. Freeman, San Francisco, 1968).
[8] P. H. Richter and M. Eigen, Biophys. Chem. **2**, 255 (1974).
[9] K. Pant, R. L. Karpel, I. Rouzina, and M. C. Williams, J. Mol. Biol. **336**, 851 (2004).
[10] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005).
[11] C. Aslangul, Europhys. Lett. **44**, 284 (1998).
[12] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, UK, 2001).
[13] S. A. Rice, in *Comprehensive Chemical Kinetics*, edited by C. H. Bamford, C. F. H. Tipper, and R. G. Compton (Elsevier, Amsterdam, 1985), Vol. 25.
[14] M. C. Williams, Optical Tweezers: Measuring Piconewton Forces. Biophysics Textbook Online: http://www.biophysics.org/btol/, 2002.
[15] J. D. McGhee and P. H. von Hippel, J. Mol. Biol. **86**, 469 (1974).
[16] K. Pant, R. L. Karpel, I. Rouzina, and M. C. Williams, J. Mol. Biol. **349**, 317 (2005).
[17] R. Metzler and J. Klafter, J. Phys. A **37**, R161 (2004).
[18] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
[19] I. M. Sokolov, J. Klafter, and A. Blumen, Phys. Today **55**(11), 48 (2002).
[20] A. Hanke and R. Metzler, Biophys. J. **85**, 167 (2003).
[21] I. M. Sokolov, J. Mai, and A. Blumen, Phys. Rev. Lett. **79**, 857 (1997).
[22] I. M. Sokolov, J. Mai, and A. Blumen, J. Lumin. **76–77**, 377 (1998).
[23] M. Lomholt, T. Ambjörnsson and R. Metzler, e-print cond-mat/0510072.
[24] A. M. Berezhkovskii, D. J. Bicout, and G. H. Weiss, J. Chem. Phys. **110**, 1112 (1999); D. J. Bicout, A. M. Berezhkovskii, and A. Szabo, *ibid.* **114**, 2293 (2001).
[25] R. D. Astumian and P. Hänggi, Phys. Today **55** (11), 33 (2002).